SC/66a/EM/3

A preliminary review of species distribution models (SDMs) applied to baleen whales

Hiroto Murase, Ari Friedlaender, Natalie Kelly, Daniel M. Palacios, Debra Palka

INTERNATIONAL WHALING COMMISSION

A preliminaly review of species distribution models (SDMs) applied to baleen whales

(Progress report of the intersessional corresponding group "Applications of species distribution models (SDMs)")

HIROTO MURASE[1], ARI FRIEDLAENDER[2, 3], NATALIE KELLY[4], DANIEL M. PALACIOS[2], DEBRA PALKA[5]

[1]*National Research Institute of Far Seas Fisheries, Fisheries Research Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa, 236-8648, Japan*
[2]*Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, 2035 Marine Science Drive, Newport, Oregon 97365, USA*
[3]*Southall Environmental Associates, Inc. 9099 Soquel Drive, Suite 8, Aptos, California 95003, USA*
[4]*CSIRO Digital Productivity and Services Flagship, Castray Esplanade, Hobart, Tasmania, 7000, Australia*
[5]*NOAA Fisheries, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, Massachusetts, 02543, USA*

ABSTRACT

A statistical model relating occurrence of a species and its environment at a certain time period is termed a species distribution model (SDM). It can be used to predict spatial distributions of the target species once the model is constructed. An intersessional correspondence group was established during IWC/SC 65b to contribute development of a guideline on the techniques and underlying assumptions of SDMs based on up-to-date and comprehensive knowledge. This review is a progress report of the group. In this manuscript, general aspects of SDMs are reviewed firstly, then SDMs applied to baleen whales are reviewed. SDMs can be categorized into four approaches based on response variables: presence/absence, presence only, presence/background and presence/pseudoabsence approaches. Abundance can be used in some models in substitution for presence. SDMs can also be categorized into four methods based on statistical models: regression, profile, machine learning and other methods. Several points such as spatial autocorrelation and collinearity need to be considered prior to modelling. SDMs have been applied to baleen whales since late 1990s. A total of 36 papers published from December 1997 to March 2015 are reviewed in this manuscript. SDMs were applied to all baleen whale species except pygmy right and Omura's whales. A total of 10 types of statistical models were used in these studies. Although the results significantly contribute to expand our knowledge on baleen whale ecology, detailed descriptions of construction and evaluation methods are needed for further consideration of the results. It is recommended that comparison among different SDMs and ensemble modelling should be pursued in future studies because it has been well documented that different models created different prediction results. It is also recommended that an appropriate guideline for parameter settings should be prepared as they are commonly applied to baleen whales.

INTRODUCTION

Spatial distribution of biological organisms is one of the fundamental information sources used for management and conservation of the target species. Classically, geographic locations of target species' occurrence are plotted on maps to understand their spatial distributions. Application of species distribution models (SDMs) (e.g. Franklin, 2009) has proliferated since 1990s in parallel with the advancement of computing power, software such as geographic information systems (GIS) and statistical techniques. Application of SDMs to baleen whales started in the late 1990s and the number of studies has been increasing in recent years. In this manuscript, a statistical model relating occurrence of a species to its environment at a certain time period is termed an SDM. Such a model can be used to predict spatial distributions of the target species once the model is constructed. An SDM is not a mechanistic model that can deal with dynamic processes of the spatial distributions but an empirical model that can incorporate observed relationships between occurrence of species and their environment at a certain time period (Palacios *et al.*, 2013). A SDM can also be called as a snapshot model, being static rather than dynamic.

Other terms such as habitat distribution model (Guisan and Zimmermann, 2000), ecological niche model (Peterson *et al.*, 2011) and resource selection function (Manly *et al.*, 2002) are sometimes used interchangeably with SDM. Habitat, niche and resource selection modelling are important and interesting topics in the field of ecology, but the definition of these terms can be bewildering and the interpretation of output from models in these contexts is sometimes difficult, especially for baleen whales as they can use a large geographic area and display complex behaviour such as migration between breeding and feeding grounds. The term "SDM" is used throughout this manuscript because the main focus is prediction of spatial distribution of target species. However, it should be noted that within the SDM framework, there is also scope for addressing ecological questions and hypotheses, but one does have to be careful about the selection and interpretation of explanatory and/predictor variables. Detailed textbooks dealing with these kinds of models are available, such as Franklin (2009) and Peterson *et al.* (2011), and these are referred throughout this manuscript. A review of application of SDMs to marine species was also published (Robinson *et al.*, 2011).

Within the Scientific Committee of the International Whaling Commission (IWC/SC), a generalized additive model (GAM) based SDM was developed in the late 1990s for the purpose of generating an abundance estimate for Antarctic minke whales (*Balaenoptera bonaerensis*) (Hedley *et al.*, 1999). Since then, SDMs have been applied to a variety of species and regions in the IWC/SC to address questions such as reasons of changes in abundance and spatial distribution of baleen whales (e.g. Beekmans *et al.*, 2010; Murase *et al.*, 2013; Williams *et al.*, 2014). Traditionally, abundance of baleen whales for the purpose of management under the Revised Management Procedure (RMP) have been estimated using a statistical design-based methods, such as the DISTANCE sampling, according to a guideline from the IWC/SC (IWC, 2012). The Sub-Committee on the RMP of the IWC/SC is currently trying to develop a guideline for model-based abundance estimation methods, mainly focusing on GAMs (Hedley and Bravington, 2014). Statistical models other than GAMs are also used as SDMs, mainly focusing on prediction of spatial distribution. The Working Group on Ecosystem Modelling (EM) of the IWC/SC recognized the necessity for the development of a guideline on the techniques and underlying assumptions of SDMs based on up-to-date and comprehensive knowledge (IWC, 2015). An intersessional correspondence group was established during IWC/SC 65b to facilitate this work. This manuscript is one of the outcomes from the group. The primary term of reference for the group was to "develop guidelines and recommendations for best practices in modelling steps". Estimation of abundance is main focus in the context of RMP while explanatory investigation on spatial distribution in relation to environment is main focus in the context of EM though the distinction is not clear-cut.

Outside of the IWC, a series of workshops on habitat modelling for marine mammals have been held at the Biennial Conference of the Society for Marine Mammalogy since 2001, and these activities led to the theme section of *Endangered Species Research* titled "Beyond Marine Mammal Habitat Modeling" (Gregr *et al.*, 2013). Several papers in the theme section are reviewed in this manuscript. A review paper of habitat modelling techniques for cetaceans was also published in *Marine Ecology Progress Series* in 2006 (Redfern *et al.*, 2006). Habitat models targeting fish were reviewed by de Kerchove *et al.* (2008). Some habitat modelling studies did not predict spatial distributions even if same statistical models as SDMs were used (e.g. Friedlaender *et al.*, 2006). Habitat modeling studies without prediction of spatial distribution are not considered fully in this manuscript.

The aim of this manuscript is three fold. Firstly, general aspects of SDMs are briefly reviewed. Secondly, applications of SDMs on baleen whales are reviewed. To accomplish the second aim, a total of 36 published papers are reviewed. Finally, preliminary recommendations are provided to develop the methods further. Geostatistical methods such as inverse distance weighting and kriging have also been used to predict spatial distributions of baleen whales. Geostatistical methods are not reviewed fully in this manuscript because spatial distribution of species are predicted based thoroughly on geographic coordinates by the methods without consideration of environmental factors.

GNERAL REVIEW OF SDMS

Data for SDMs

Data on occurrence/abundance of species (response variable)

To construct SDMs, data on geographic locations of observed species (i.e., response variable) and environmental data at the time of observations (i.e., explanatory or predictor variables) is required. Geographic locations and timestamps of the observations are fundamental data for SDMs. There are two basic types of response variables: (i) occurrence (when and where individuals and groups were detected during a survey) and (ii) abundance (i.e., number of individuals detected). Abundance (or, equally, density) data can be considered as an extension of occurrence data. Probability of occurrence can be estimated if occurrence data are used, while abundance can be estimated if abundance data are used.

If occurrence or abundance data are obtained from a target area based on survey activity recording when and where search effort was undertaken (regardless of whether such a survey is based on a predetermined statistical survey design, or an opportunistic, non-random design), both occurrence (i.e. presence) and absence data can be obtained. In contrast, data such as telemetry and catch without effort data can only provide occurrence data. SDMs can be categorized by four approaches based on available response variables: (i) presence/absence, (ii) presence only, (iii) presence/background, and (iv) presence/pseudo-absence. Background data implies the existence of co-collected environmental data in a target area. Pseudo-absence implies existence of environmental data in a target area, but environmental data at occurrence positions are excluded. Presence/background and presence/pseudo-absence approaches were developed to compensate cases when absence data are not available (Phillips and Elith (2013) or Hastie and Fithian (2013) for further discussion).

Environmental data (explanatory/predictor variables)
There are three categories of environmental data which can be used in SDMs targeting marine species: (i) topographic variables, (ii) physical and chemical oceanographic variables, and (iii) biological variables. One of the main topographic variables used is bottom depth. Several variables can be calculated using bottom depth data such as slope, bottom complexity and isobath. Application of topographic variables to studies on mobile vertebrate was reviewed by Bouchet *et al.* (2014). Bottom substrate can also be considered as a topographic variable. Distance from topographic features such coastline and shelf break can be used as explanatory or predictor variables, as these can be assumed to be potentially informative aliases or proxies for biological/ecological processes influencing the distribution of target species. Examples of physical and chemical oceanographic variables are temperature, salinity, sea surface height, current speed and direction, frontal boundary, mixed layer depth, dissolved chemicals and sea ice concentration. Some of these variables were reviewed by Hobday and Hartong (2014). Chlorophyll-*a* concentrations and prey densities are commonly used biological variables, but information on predators and competitors can also be used as variables if available. Some of the satellite-derived sea surface variables, such as temperate, sea-surface height, current and sea ice concentrations, are available through a number of agencies without cost for non-commercial use (e.g. National Snow and Ice Data Center: www.nsidc.org). Satellite data are useful explanatory/predictor variables for baleen whales as spatial coverage of most of them are global, often with frequent re-sampling over days, weeks or months. Geographic coordinates (e.g. longitude and latitude) might be used as explanatory variables to take account of spatial variations that can not be captured by environmental factors alone, but, of course, such relationships are of little or no use in testing ecological hypotheses if there are no inferences upon the actual underlying ecological processes at work across geographical space. As well, spatial coordinates are critical for analyses that can account for spatial autocorrelation amongst environmental variables (e.g. Mantel’s tests) when considering their effect on distribution patterns of species (e.g. Friedlaender *et al.*, 2006; see below for more information).

Spatial and temporal scales of data
It is ideal that both response and explanatory variables are recorded simultaneously. However, necessary environmental data might not be recorded at the time of observation of species because of availability of measurement instruments or logistical constraints. Furthermore, environmental data covering the entire target area is required to predict, via interpolation and/or extrapolation, spatial distribution of target species. In this way, environmental data are generally more useful than geographical coordinates for interpolation and/or extrapolation if one assumes relationships between environmental covariates and densities of target species remain stable over the study area. Principally, selection of spatial and temporal scales of environmental data should be based on sound ecological reasoning. However, availability of data could limit such a selection. For instance, if the coarsest spatial and temporal scales of environmental data are 5×5 km grid cell and month, respectively, resolutions of other data with finer resolutions should be resampled to match the coarsest resolutions.

Types of SDMs
Several types of statistical models are used as SDMs. These models can be classified into three methods according to Hijmans and Elith (2015): (i) regression method, (ii) profile method and (iii) machine learning method. However, some models can not fit to this classification and these are grouped under other methods in this manuscript. Ensembles of models are also mentioned here briefly. As mentioned in the previous section, SDMs can be classified based on type of response variables. Classification of some major SDMs by statistical models and types of response variables is summarized in Table 1. The presence/pseudo-absence approach can be considered as a special case of the presence/absence approach. Statistical models which can deal with presence/absence data can basically deal with presence/pseudo-absence data.

Regression method
A general linear model is a fundamental regression method that assumes a linear relationship between response and explanatory variables. However, relationship between spatial distribution of species and their environment could be more complex. A generalized linear model (GLM) (McCullagh and Nelder 1989) and generalized additive model (GAM) (Wood, 2006) are commonly used as SDMs to deal with such complexities. GLM and GAM require presence/absence data. Abundance instead of presence data can also be used in these models. GAMs are particularly favoured owing to their flexibility for allowing relationships (existing or, admittedly, spurious) between response and independent variables to drive the parameterisation process, and not pre-conceived constraints on model format.

Profile method
In the profile method, distinctive environmental conditions at locations of presence are identified through some values such as means and ranges. The profile method is best suited to presence-only data. A habitat suitability indices (HSI) (Verner *et al.*, 1986), bioclimatic analysis and prediction system (BIOCLIM) (Busby, 1991) and DOMAIN (Carpenter *et al.*, 1993) are used as SDMs. Relative environmental suitability model (RES) (Kaschner *et al.*, 2006) can be considered as a type of HSI.

Machine learning method
In the machine learning method, species distributions in relation to their environment are determined based on certain rules. Basically, presence/absence data are required in the machine learning methods. Decision-tree based models such as Classification and Regression Tree (CART) (Breiman *et al.*, 1984) random forest (RF) (Breiman, 2001) and boosted regression tree (BRT) (Elith *et al.*, 2008) are categorized as machine learning methods (see also De'ath (2007) and De'ath and Fabricius (2000) for further details). Other models such as neural network (NN) (Pearson *et al.*, 2002) and support vector machine (Guo *et al.*, 2005) are also used as SDMs. A maximum entropy-based method called MaxEnt (Elith *et al.*, 2011) is specifically designed to use presence/background data. Because software for MaxEnt is available (https://www.cs.princeton.edu/~schapire/maxent/), and one can run it without programing once a data set is prepared, MaxEnt is frequently used as SDMs (however, see Kramer-Schadt *et al.* (2013) and Yackulic *et al.* (2013) for further discussion on potential problems and biases with using the MaxEnt approach).

Other methods
Hierarchical Bayesian model (HBM) is used as a SDM to deal explicitly with heterogeneity in detection of species (Royle and Dorazio, 2008). Other advantages of HBM are as follows: allowing for the explicit propagation of uncertainty, and for several sub-models to be seamlessly integrated; allowing for the estimation of abundance (through Distance sampling) simultaneously with the estimation of association with environmental variables as separate sub-models (see Pardo *et al.*, 2015 for more details). Species distribution is characterized to take account of contrasts between environmental conditions at presence of species and the background in Ecological Niche Factor Analysis (ENFA) (Hirzel *et al.*, 2002).

Ensemble model
It was documented that different SDMs applied to the same data sets created different prediction results (Elith *et al.*, 2006; Segurado and Araújo, 2004). Ensemble modelling (Araújo and New, 2007) is applied to deal with errors and uncertainties of each SDM. Four ensemble methods are proposed by Araújo and New (2007): (i) bounding box or generating a consensus forecast for small ensemble size; (ii) showing number of models forecasting presence using histogram; (iii) showing probability density function of likelihood of species presence for large ensembles and (iv) measuring central tendency (e.g. mean and median).

Points to be considered before modelling

Sampling of response variable
It is desirable that the response variable (occurrence of species) be randomly sampled from a target area based on robust statistical design, or, if this is too strong an assumption, at least randomly sampled in relation to the environmental covariates. Even if data are not sampled randomly, it is ideal to ensure broad coverage, perhaps via a systematic survey design, throughout a target area, particularly if distributions of a target species have, in that area, have previously not been studied. Obviously, the survey costs, logistics, and requirements for minimum useful sample sizes will also influence survey design. However, commercial catch and telemetry data could violate such assumptions as the nature of these data sets is basically non-random and, spatial and temporal coverages are limited. It is necessary to check whether the data are potentially biased, both spatially and temporally, before the modelling. Careful interpretations of the outcomes of SDMs are necessary if there are such biases. It is desirable to use both presence and absence data in SDMs whenever these data are available as it is expected that these data contained much information on environmental conditions that constrain presence of species in comparison with presence only data.

Reliability of response variable
Occurrence of species must be recorded correctly. However, it might not be in some cases such as miss-identification of species (Conn *et al.*, 2013) and imperfect detection (Laake *et al.*, 2008). Appropriate analytical treatment is necessary if either are suspected. Much has been written about biases introduced via heterogeneity in detection probabilities in fauna and flora surveys (e.g., Borchers *et al.* (2006), Ramsey and Harrison (2004) and Thomson et al. (2012) to name but a few examples).

Spatial autocorrelation of response variable

Spatial autocorrelation (SAC) in the context of SDMs was reviewed by Dormann *et al.* (2007). SAC occurs when data sampled in close proximity are not independent from each other. In a general sense, samples that are too similar, in this example, because they are adjacent in space and time, will yield falsely low variance estimates. Independence among data, which is assumed in standard statistical models, is violated if SAC exists and it can lead to type I error. Existence of SAC can be checked by some indices such as Moran's I and Geary's C, or explored using geostatistical or mixed-effects modelling. Several statistical models which can deal with SAC are available (Table 2). It is likely that telemetry data are suspect to SAC as the data obtained continuously from tagged species.

Ecological validity of explanatory variables

Ideally, selection of explanatory variables should be at least broadly based on ecological reasoning, which can be obtained qualitatively as expert knowledge. However, the selection would be arbitrary if such information is not available before the modelling. In a predictive modelling context, there may be little or no desire to capture information about actual ecological/biological processes—but if it is captured in models, this can probably only help. With predictive modelling, the aim is produce models optimised for accurate predictions within the bounds of the system of interest (i.e., with little to no extrapolation). Ideally, explanatory variables are cheap to collect, and plentiful, as the more information will lead to more accurate predictions—any relation they have to the variable of interest (here, the presence or densities of species) can be purely coincidental, just as long as it is useful for accurate prediction. For more thorough discussions on the differences between predictive and explanatory modelling, see Mac Nally (2000) and Shmueli (2010).

Collinearity among explanatory variables

Collinearity in the context of SDMs was reviewed by Dormann *et al.* (2012). Collinearity implies that some of explanatory variables, especially in regression methods, are related. If collinearly related variables exist in a model, explanatory power of one of the collinearly related variables might be reduced and/or the model can be unstable. Collinearity can be serious problem when a selected model is used for prediction where structure of collinearity is unknown. Existence of collinearity can be checked using some indices such as variance inflation factor (VIF). Several methods which can be applied before modelling or during modelling are available to deal with collinearity as reviewed in Dormann *et al.* (2012). Hierarchical partitioning is one method that assists in teasing out the effects of collinearity with the aim of identifying potentially important explanatory variables (Mac Nally 2000).

Model Evaluation

Model evaluation methods for presence/absence data was reviewed by Fielding and Bell (1997). Area under the curve (AUC) of receiver operating characteristic curve (ROC) is one commonly used method. For abundance data, predictions and overall model performance were compared using explained deviance, average squared prediction error (ASPE), and ratios of observed to predicted densities to identify the best models (Forney *et al.*, 2012).

REVIEW OF APPLICATIONS OF SDMS TO BALEEN WHALES

A total of 36 published in scientific journals from December 1997 to March 2015 are considered to review applications of SDMs to baleen whales.

Target species, regions and areas

Target species, regions and areas considered in the published papers are summarized in Table 3. SDMs were applied to all baleen whales except pygmy right (*Caperea marginata*) and Omura's (*B. omurai*) whales. SDMs were applied to a wide varieties of regions and areas, but there are few applications in the Indian Ocean and South Pacific. No study was conducted in the South Atlantic. Some papers dealt with multi-species and/or multi-regions.

Data

Response variable

Response variables used in the published papers are summarized in Table 4. Data sets obtained by dedicated sighting surveys were used in 17 papers followed by data sets obtained by opportunistic sighting surveys (13 papers). In this review, surveys conducted based on the DISTANCE sampling method were defined as dedicated surveys. Catch data and published data were also used in SDMs. Count (number of animals or abundance) data sets were used in 18 papers followed by presence only (11 papers) and presence/absence (7 papers) data sets.

Explanatory variables

Explanatory variables used in the published papers are summarized in Table 5. A total of 27 types of explanatory variables were used in the published papers. Bottom depth was the most commonly used variable in the SDMs (29 papers) followed by sea surface temperature (SST) (23 papers), seafloor slope (16 papers), surface chlorophyll-*a* concentrations (15 papers), longitude (12 papers), latitude (11 papers) and distance to shore (11 papers). Other variables were used in less than 11 papers.

Statistical models used as SDMs

Statistical models used in the published papers are summarized in Table 6. Only one paper considered two different statistical models. Other papers used only one statistical model per analysis. GAMs including one paper that used mixed-effects GAM (GAMM) were the most used models (17 papers) followed by MaxEnt (7 papers). Other models (BRT, CART, ENFA, GLM, HBM, logistic regression and RES) were used in 3 or fewer papers. One paper used a geostatistical method, kernel density smoothing, to predict spatial distribution of baleen whales (Laidre *et al.*, 2010). Forney *et al.* (2012) and Víkingsson *et al.* (2015) used GAMs as SDMs, but the spatial distributions were predicted by geostatistical methods, namely inverse distance weighting (IDW) and kriging, respectively. SACs were considered in 5 papers, using various methods. Collinearity was considered in 3 papers. Resolutions of grid cells used for spatial distributions were varied among papers (approximately from 1 to 50 km).

Parameter setting for GAMs

Some of parameter setting for GAMs are reviewed here as GAMs have been used frequently as SDMs for baleen whales. Common parameter settings for GAMs were summarized in Table 7. Some of the settings are specific to the mgcv package (Wood, 2006) of the R software (R Development Core Team, 2015): dimension parameter (k), and gamma. Dimension parameter (k) sets the upper limit on the degrees of freedom. A constant multiplier to inflate the model degrees of freedom in the GCV (generalized cross validation) or UBRE/AIC (un-biased risk estimator/Akaike information criterion) is called as “gamma”.

Response variables based on count data often suffer from overdispersion and zero inflation (i.e., more variation than can be handed by a Poisson distribution). To account for this, GAMs (and GLMs) can use quasi-Poisson, negative binomial or Tweedie distributions.

Some GAM parameter settings were not reported in several papers but it was likely that they used default settings (which, in general, should not lead to completely incorrect outcomes as Simon Wood, the author of mgcv, worked hard to set the defaults to represent an overall reasonable compromise across manifold complex considerations). Model selections were based on either GCV, AIC, BIC (Bayesian information criterion) or REML (restricted maximum likelihood).

Whilst GAMs allow flexibility, and for features of the data to drive model parameterisation, it is often the case that relationships between explanatory and response variables can be either under or over-described or ‘smoothed’. For example, a GAM might indicate that density of some species has three or more ‘maxima’ over a range of a given environmental covariate. If the covariate represents a biological gradient, is it likely that a particular species has several optimal values over such a variable? Therefore, care does need to be taken in pre-defining (or not) assumptions about ‘wiggliness’ of densities along gradients.

DISCUSSION

A wide variety of statistical models as SDMs have been applied to most of the baleen whale species known worldwide. The results significantly contribute to expand our knowledge on their ecology and behaviour. It seems that applications of SDMs to baleen whales up to now are exploratory as most cetacean scientists are still in the capacity building stage in this field. Most papers only applied a statistical model and comparison of results from different models applying to a data set has not been conducted fully. As mentioned above, it was reported that SDMs applied to same data sets created different prediction results (Elith *et al.*, 2006; Segurado and Araújo, 2004). Therefore, it is recommended that comparison of results of different models should be conducted in future studies. Furthermore, ensemble modelling should also be attempted to deal with errors and uncertainties of each SDM. However, it should be noted these studies only considered models to deal with probability of occurrence. A study was conducted to compare results among SDMs which use count data of seabird as a response variable (Renner *et al.*, 2013). Renner *et al.* (2013) also attempted ensemble modelling using count data. Oppel *et al.* (2012) also compared and attempted ensemble modelling using seabird data. Such attempts should also be conducted using baleen whale data.

Construction methods of SDMs applied to baleen whales were not consistent. For instance, SACs and collinearity were not considered in most of papers, or, at least, not directly reported in such studies. It is recommended to consider at least these two factors when constructing SDMs for baleen whales as they might have significant effect on the results of modelling, dependent on model objectives.

GAMs were the most commonly used SDMs applied to baleen whales, probably because an established method in the cetacean literature (Forney 2000; Hedley *et al.*, 1999; Palka 1995) is available since the 1990s. There are a number of settings for GAM modelling but there is no consensus method or documentation to help set them although Forney *et al.* (2012) provided a good summary. The setting will be different for different data sets, but some guidelines are required to narrow down the choices. It is recommended that the point should be considered in on-going work of the RMP sub-committee on development of a guideline for model based abundance estimation method. Setting for other models such as MaxEnt should also be considered in the future work.

REFERENCES

Ainley, D., Jongsomjit, D., Ballard, G., Thiele, D., Fraser, W., Tynan, C. 2012. Modeling the relationship of Antarctic minke whales to major ocean boundaries. *Polar Biol.* 35: 281-90.

Araújo, M.B., New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22: 42-7.

Ballard, G., Jongsomjit, D., Veloz, S.D., Ainley, D.G. 2012. Coexistence of mesopredators in an intact polar ocean ecosystem: The basis for defining a Ross Sea marine protected area. *Biol. Conserv* 156: 72-82.

Becker, E.A., Foley, D.G., Forney, K.A., Barlow, J., Redfern, J.V., Gentemann, C.L. 2012. Forecasting cetacean abundance patterns to enhance management decisions. *Endanger. Species Res.* 16: 97-112.

Beekmans, B.W.P.M., Forcada, J.F., Murphy, E.J., de Baar, H.J.W., Bathmann, U.V., Fleming, A.H. 2010. Generalised additive models to investigate environmental drivers of Antarctic minke whale (*Balaenoptera bonaerensis*) spatial density in austral summer. *J. Cetacean Res. Manage.* 11: 115-29.

Best, B.D., Halpin, P.N., Read, A.J., Fujioka, E., Good, C.P., LaBrecque, E.A., Schick, R.S., .J., R.J., Hazen, L.J., Qian, S.S., L., P.D., P., G.L., McLellan, W.A. 2012. Online cetacean habitat modeling system for the US east coast and Gulf of Mexico. *Endanger. Species Res.* 18: 1-15.

Bombosch, A., Zitterbart, D.P., Van Opzeeland, I., Frickenhaus, S., Burkhardt, E., Wisz, M.S., Boebel, O. 2014. Predictive habitat modelling of humpback (*Megaptera novaeangliae*) and Antarctic minke (*Balaenoptera bonaerensis*) whales in the Southern Ocean as a planning tool for seismic surveys. *Deep Sea Res. I* 91: 101-14.

Borchers, D. L., Laake, J. L., Southwell, C. and Paxton, C. G. M. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics* 62: 372-78.

Bouchet, P.J., Meeuwig, J.J., Salgado Kent, C.P., Letessier, T.B., Jenner, C.K., 2014. Topographic determinants of mobile vertebrate predator hotspots: current knowledge and future directions. *Biol. Rev.* doi:10.1111/brv.12130

Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1984. *Classification and regression trees*. Chapman & Hall/CRC London. 358pp.

Busby, J.R. 1991. BIOCLIM - A Bioclimatic Analysis and Prediction System. pp.64-8 *In*: C.R. Margules, M.P. Austin (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Canberra

Carpenter, G., Gillison, A.N., Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2: 667-80.

Conn, P. B., McClintock, B. T., Cameron, M. F., Johnson, D. S., Moreland, E. E. and Boveng, P. L. 2013. Accommodating species identification errors in transect surveys. *Ecology.* 94: 2607-18.

Corkeron, P.J., Minton, G., Collins, T., Findlay, K., Willson, A., Baldwin, R. 2011. Spatial models of sparse data to inform cetacean conservation planning: an example from Oman. *Endanger. Species Res.* 15: 39-52.

Cotté, C., Guinet, C., Taupier-Letage, I., Mate, B., Petiau, E. 2009. Scale-dependent habitat use by a large free-ranging predator, the Mediterranean fin whale. *Deep Sea Res. I* 56: 801-11.

Dalla Rosa, L., Ford, J.K.B., Trites, A.W. 2012. Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Cont. Shelf Res.* 36: 89-104.

De'ath, G. and Fabricius, K. E. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecolgy* 81: 3178-92.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88: 243-251.

De Kerckhove, D.T., Smokorowski, K.E., Randall, R.G. 2008. A Primer on Fish Habitat Models. *Can. Tech. Rep. Fish. Aquat. Sci.* 2817: 1-65

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore,

A.K., Zurell, D., Lautenbach, S. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*: 27-46.
Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ralf, O., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609-28.
Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-51.
Elith, J., Leathwick, J.R., Hastie, T. 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77: 802-13.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.
Fielding, A.H., Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.
Forney, K.A. 2000. Environmental models of cetacean abundance: reducing uncertainty in population trends. *Conserv. Biol.* 14: 1271-86.
Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J.V., Barlow, J., Vilchis, I.L., Ballance, L.T. 2012. Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endanger. Species Res.* 16: 113-33.
Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge.
Friedlaender, A.S., Halpin, P.N., Qian, S.S., Lawson, G.L., Wiebe, P.H., Thiele, D., Read, A.J. 2006. Whale distribution in relation to prey abundance and oceanographic processes in shelf waters of the Western Antarctic Peninsula. *Mar. Ecol. Prog. Ser.* 317: 297-310.
Friedlaender, A.S., Johnston, D.W., Fraser, W.R., Burns, J., Patrick N, H., Costa, D.P. 2011. Ecological niche modeling of sympatric krill predators around Marguerite Bay, Western Antarctic Peninsula. *Deep Sea Res. II* 58: 1729-40.
Gregr, E.J. 2011. Insights into North Pacific right whale Eubalaena japonica habitat from historic whaling records. *Endanger. Species Res.* 15: 223-39.
Gregr, E.J., Baumgartner, M.F., Laidre, K.L., Palacios, D.M. 2013. Marine mammal habitat models come of age: the emergence of ecological and management relevance. *Endanger. Species Res.* 22: 205-12.
Gregr, E.J., Trites, A.W. 2001. Predictions of critical habitat for five whale species in the waters of coastal British Columbia. *Can. J. Fish. Aquat. Sci.* 58: 1265-85.
Guisan, A., Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-86.
Guo, Q., Kelly, M., Graham, C.H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecol. Model.* 182: 75-90.
Hamazaki, T. 2002. Spatiotemporal prediction models of cetacean habitats in the Mid-Western North Atlantic Ocean (from Cape Hatteras, North Carolina, U.S.A. to Nova Scotia, Canada). *Mar. Mamm. Sci.* 18: 920-39.
Hastie, T. and Fithian, W. 2013. Inference from presence-only data; the ongoing controversy. *Ecography* 36: 864-867.
Hedley, S., Bravington, M. 2014. Comments on design-based and model-based abundance estimates for the RMP and other contexts. Paper SC/65b/RMP11presented to the 65b IWC Scientific Committee, May 2014. (unpublished). 9pp.
Hedley, S.L., Buckland, S.T., Borchers, D.L. 1999. Spatial modelling from line transect data. *J. Cetacean Res. Manage.* 1: 255-64.
Hijmans, R.J., Elith, J., 2015. Species distribution modeling with R (March 14, 2015), 78pp.
Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-36.
Hobday, A.J., Hartog, J.R. 2014. Derived ocean features for dynamic ocean management. *Oceanography* 27: 134-45.
IWC 2012. Requirements and guidelines for conducting surveys and analysing data within the revised management scheme. *J. Cetacean Res. Manage.* 13 (Suppl.): 509-17.
IWC 2015. Annex K1 Report of the Working Group on Ecosystem Modelling. *J. Cetacean Res. Manage.* 16 (suppl.) (in press).

Kaschner, K., Watson, R., Trites, A.W., Pauly, D. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Mar. Ecol. Prog. Ser.* 316: 285-310.
Keller, C.A., Garrison, L., Baumstark, R., Ward-Geiger, L.I., Hines, E. 2012. Application of a habitat model to define calving habitat of the North Atlantic right whale in the southeastern United States. *Endanger. Species Res.* 18: 73-87.
Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19: 1366-79.
Laake, J., Guenzel, R. J., Bengtson, J. L., Boveng, P., Cameron, M. and Hanson, M. B. 2008. Coping with variation in aerial survey protocol for line-transect sampling. *Wildlife Res.* 35: 289-98.
Laidre, K.L., Heide-Jorgensen, M.P., Heagerty, P., Cossio, A., Bergstrom, B., Simon, M. 2010. Spatial associations between large baleen whales and their prey in West Greenland. *Mar. Ecol. Prog. Ser.* 402: 269-84.
Laran, S., Gannier, A. 2008. Spatial and temporal prediction of fin whale distribution in the northwestern Mediterranean Sea. *ICES J. Mar. Sci.* 65: 1260-9.
Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of 'predictive' and 'explanatory' models. *Biodivers. Conserv.* 9: 655-71.
Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P. 2002. *Resource selection by animals: statistical design and analysis of field studies*, 2nd edition ed. Kluwer Academic Publishers, Dordrechut.
McCullagh, P. and J. A. Nelder 1989. *Generalized Linear Models*. Oxford, Chapman and Hall.
Moses, E., Finn, J.T. 1997. Using Geographic Information Systems to predict North Atlantic right whale (*Eubalaena glacialis*) habitat. *J. North. Atl. Fish. Sci.* 22: 37-46.
Murase, H., Hakamada, T., Matsuoka, K., Nishiwaki, S., Inagake, D., Okazaki, M., Tojo, N., Kitakado, T. 2014. Distribution of sei whales (*Balaenoptera borealis*) in the subarctic–subtropical transition area of the western North Pacific in relation to oceanic fronts. *Deep Sea Res. II* 107: 22-8.
Murase, H., Kitakado, T., Hakamada, T., Matsuoka, K., Nishiwaki, S., Naganobu, M. 2013. Spatial distribution of Antarctic minke whales (*Balaenoptera bonaerensis*) in relation to spatial distributions of krill in the Ross Sea, Antarctica. *Fish. Oceanogr.* 22: 154-73.
Okamura, H., Matsuoka, K., Hakamada, T., Okazaki, M., Miyashita, T. 2001. Spatial and temporal structure of the western North Pacific minke whale distribution inferred from JARPN sightings data. *J. Cetacean Res. Manage.* 3: 193-200.
Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol. Conserv.* 156: 94-104.
Palacios, D.M., Baumgartner, M.F., Laidre, K.L., Gregr, E.J. 2013. Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. *Endanger. Species Res.* 22: 191-203.
Palka, D. 1995. Influences on spatial patterns of Gulf of Maine harbor porpoises. pp.69-75 In: L.W. Arnoldus Schytte Blix, U. Øyvind (Eds.), Developments in Marine Biology. Elsevier Science
Panigada, S., Zanardelli, M., MacKenzie, M., Donovan, C., Melin, F., Hammond, P.S. 2008. Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sens. Environ.* 112: 3400-12.
Pardo, M.A., Gerrodette, T., Beier, E., Gendron, D., Forney, K.A., Chivers, S.J., Barlow, J., Palacios, D.M. 2015. Inferring Cetacean Population Densities from the Absolute Dynamic Topography of the Ocean in a Hierarchical Bayesian Framework. *PLoS ONE* 10: e0120727.
Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecol. Model.* 154: 289-300.
Pendleton, D.E., Sullivan, P.J., Brown, M.W., Cole, T.V.N., Good , C.P., Mayo, C.A., Monger, B.C., Phillips, S., Record, N.R., Pershing, A. 2012. Weekly predictions of North Atlantic right whale Eubalaena glacialis habitat reveal influence of prey abundance and seasonality of habitat preferences. *Endanger. Species Res.* 18: 147-61.
Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
Phillips, S. J. and Elith, J. 2013. On estimating probability of presence from use-availability or presence-background data. *Ecology* 94: 1409-19.

R Development Core Team 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ramsey, F. L. and Harrison, K. 2004. A closer look at detectability. *Environ. Ecol. Stat.* 11: 73-84.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L., Werner, F. 2006. Techniques for cetacean-habitat modeling. *Mar. Ecol. Prog. Ser.* 310: 271-95.

Renner, M., Parrish, J.K., Piatt, J.F., Kuletz, K.J., Edwards, A.E., Hunt, G.L.J. 2013. Modeled distribution and abundance of a pelagic seabird reveal trends in relation to fisheries. *Mar. Ecol. Prog. Ser.* 484: 259-77.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20: 789-802.

Royle, J.A., Dorazio, R.M. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Elsevier Science. 444pp.

Sasaki, H., Murase, H., Kiwada, H., Matsuoka, K., Mitani, Y., Saitoh, S. 2013. Habitat differentiation between sei (*Balaenoptera borealis*) and Bryde's whales (*B.* br*ydei*) *in* the western North Pacific. *Fish. Oceanogr.* 22: 496-508.

Segurado, P., Araújo, M.B. 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* 31: 1555-68.

Shmueli, G. 2010. To Explain or to Predict? *Stat. Sci* 25: 289-310.

Skern-Mauritzen, M., Johannesen, E., Bjørge, A., Øien, N. 2011. Baleen whale distributions and prey associations in the Barents Sea. *Mar. Ecol. Prog. Ser.* 426: 289-301.

Skov, H., Gunnlaugsson, T., Budgell, W.P., Horne, J., Nøttestad, L., Olsen, E., Søiland, H., Víkingsson, G., Waring, G. 2008. Small-scale spatial variability of sperm and sei whales in relation to oceanographic and topographic features along the Mid-Atlantic Ridge. *Deep Sea Res. II* 55: 254-68.

Smith, J.N., Grantham, H.S., Gales, N., Double, M.C., Noad, M.J., Paton, D. 2012. Identification of humpback whale breeding and calving habitat in the Great Barrier Reef. *Mar. Ecol. Prog. Ser.* 447: 259-72.

Thomson, J. A., Cooper, A. B., Burkholder, D. A., Heithaus, M. R. and Dill, L. M. 2012. Heterogeneous patterns of availability for detection during visual surveys: spatiotemporal variation in sea turtle dive-surfacing behaviour on a feeding ground. *Methods Ecol. Evol.* 3: 378-387.

Torres, L.G., Smith, T.D., Sutton, P., MacDiarmid, A., Bannister, J., Miyashita, T. 2013. From exploitation to conservation: habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Diversity and Distributions* 19: 1138-52.

Víkingsson, G.A., Heide-Jørgensen, M.P. 2015. First indications of autumn migration routes and destination of common minke whales tracked by satellite in the North Atlantic during 2001–2011. *Mar. Mamm. Sci.* 31: 376-85.

Víkingsson, G.A., Pike, D.G., Schleimer, A., Valdimarsson, H., Gunnlaugsson, T., Silva, T., Elvarsson, B.Þ., Mikkelsen, B., Öien, N., Desportes, G., Bogason, V., Hammond, P. 2015. Distribution, abundance and feeding ecology of baleen whales in Icelandic waters: have recent environmental changes had an effect? *Frontiers in Ecology and Evolution* 3.

Verner, J., Morrison, M.L., Ralph, C.J., Pacific Southwest, F., Range Experiment, S., Wildlife, S., San Francisco Bay Area, C. 1986. *Wildlife 2000 : modeling habitat relationships of terrestrial vertebrates : based on an international symposium held at Stanford Sierra Camp, Fallen Leaf Lake, California, 7-11 October 1984*. University of Wisconsin Press, Madison, Wis.

Wheeler, B., Gilbert, M., Rowe, S. 2012. Definition of critical summer and fall habitat for bowhead whales in the eastern Canadian Arctic. *Endanger. Species Res.* 17: 1-16.

Williams, R., Hedley, S.L., Branch, T.A., Bravington, M.V., Zerbini, A.N., Findlay, K.P. 2011. Chilean blue whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. *Conservation biology : the journal of the Society for Conservation Biology* 25: 526-35.

Williams, R., Hedley, S.L., Hammond, P.S. 2006. Modeling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecol. Soc.* 11.

Williams, R., Kelly, N., Boebel, O., Friedlaender, A.S., Herr, H., Kock, K.H., Lehnert, L.S., Maksym, T., Roberts, J., Scheidat, M., Siebert, U., Brierley, A.S. 2014. Counting whales in a challenging, changing environment. *Scientific Reports* 4.

Wood, S., N. 2006. *Generalized additive models: an introduction with R*. Chapman & Hall/CRC, Boca Raton.

Yackulic, C. B., Chandler, R., Zipkin, E. F.. Royle, J. A., Nichols, J. D., Grant, E. H. C. and Veran, S. 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4: 236-43.

Table 1. Classification of species distribution models (SDMs) based of modelling methods and response variables.

Method	Response variable		
	Presence/absence	Presence/background	Presence only
Regression	GLM GAM	-	-
Profile	-	-	HSI BIOCLIM DOMAIN
Machine learning	CART BRT RF ANN SVM	MaxEnt	-
Other	HBM	ENFA	-

Table 2. Some examples of statistical models which can deal with spatial autocorrelation (SAC) and other correlation structures.

Name of model	Abbreviation
Autocovariate models	-
Spatial eigenvector mapping	SEVM
Generalized least squares	GLS
Conditional autoregressive model	CAR
Simultaneous autoregressive model	SAR
Generalized linear mixed model	GLMM
Generalized estimating equations	GEE

Table 3. Target species, regions and areas of species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to March 2015 are considered. It should be noted that some papers dealt multi-species and/or multi-regions.

Species			Region	Area	Reference
Common name	Scientific name				
Bowhead whale	*Balaena*	*mysticetus*	Arctic	Eastern Canada	Wheeler *et al.* (2012)
North Atlantic right whale	*Eubalaena*	*glacialis*	North Atlantic	US east coast Gulf of Mexico	Best *et al.* (2012)
			North Atlantic	-	Kaschner *et al.* (2006)
			North Atlantic	Coast of Florida and Georgia	Keller *et al.* (2012)
			North Atlantic	Nova Scotian Shelf	Moses and Finn (1997)
			North Atlantic	Cape Cod Bay	Pendleton *et al.* (2012)
North Pacific right whale	*Eubalaena*	*japonica*	North Pacific	-	Gregr (2011)
Southern right whale	*Eubalaena*	*australis*	Southern Ocean	Australasian region	Torres *et al.* (2013)
Gray whale	*Eschrichtius*	*robustus*	North Pacific	-	Kaschner *et al.* (2006)
Minke whale	*Balaenoptera*	*acutorostrata*	North Atlantic	Mid-western North Atlantic	Hamazaki (2002)
			North Pacific	Western North Pacific	Okamura *et al.* (2001)
			Arctic	Barents Sea	Skern-Mauritzen *et al.* (2011)
Antarctic minke whale	*Balaenoptera*	*bonaerensis*	Antarctic	140°E-35°W	Ainley *et al.* (2012)
			Antarctic	Ross Sea	Ballard *et al.* (2012)
			Antarctic	Circumpolar	Beekmans *et al.* (2010)
			Antarctic	60°E-60°W	Bombosch *et al.* (2014)
			Antarctic	Marguerite Bay	Friedlaender *et al.* (2011)
			Antarctic	0°-40°E	Hedley *et al.* (1999)
			Antarctic	Circumpolar	Kaschner *et al.* (2006)
			Antarctic	Ross Sea	Murase *et al.* (2013)
			Antarctic	65°W-55°W	Williams *et al.* (2006)
			Antarctic	Weddell Sea	Williams *et al.* (2014)
Sei whale	*Balaenoptera*	*borealis*	North Pacific	British Columbia	Gregr and Trites (2001)
			North Pacific	Western North Pacific	Murase *et al.* (2014)
			North Pacific	Western North Pacific	Sasaki *et al.* (2013)
			North Atlantic	Mid-Atlantic Ridge	Skov *et al.* (2008)
Bryde's whale	*Balaenoptera*	*edeni*	Tropical Pacific	Eastern tropical Pacific	Forney *et al.* (2012)
			North Pacific	Western North Pacific	Sasaki *et al.* (2013)
Blue whale	*Balaenoptera*	*musculus*	Tropical Pacific	Eastern tropical Pacific	Forney *et al.* (2012)
			North Pacific	California Current System	Forney *et al.* (2012)
			North Pacific	British Columbia	Gregr and Trites (2001)
			South Pacific	Coast of Chile	Williams *et al.* (2011)
			Pacific Ocean	East Pacific	Pardo *et al.* (2015)
Fin whale	*Balaenoptera*	*physalus*	North Pacific	California Current System	Becker *et al.* (2012)
			Mediterranean	Western Mediterranean Sea	Cotté *et al.* (2009)
			North Pacific	California Current System	Forney *et al.* (2012)
			North Pacific	British Columbia	Gregr and Trites (2001)
			North Atlantic	Mid-western North Atlantic	Hamazaki (2002)
			Arctic	West coast of Greenland	Laidre *et al.* (2010)
			Mediterranean	Northwestern Mediterranean Sea	Laran and Gannier (2008)
			Mediterranean	Pelagos Sanctuary	Panigada *et al.* (2008)
			Arctic	Barents Sea	Skern-Mauritzen *et al.* (2011)
			Antarctic	65°W-55°W	Williams *et al.* (2006)
			North Atlantic	Island	Víkingsson *et al.* (2015)
Humpback whale	*Megaptera*	*novaeangliae*	North Atlantic	US east coast Gulf of Mexico	Best *et al.* (2012)
			Antarctic	Circumpolar	Bombosch *et al.* (2014)
			Indian Ocean	Arabian Sea off Oman	Corkeron *et al.* (2011)
			North Pacific	British Columbia	Dalla Rosa *et al.* (2012)
			North Pacific	California Current System	Forney *et al.* (2012)
			Antarctic	Marguerite Bay	Friedlaender *et al.* (2011)
			North Pacific	British Columbia	Gregr and Trites (2001)
			Arctic	West coast of Greenland	Laidre *et al.* (2010)
			Arctic	Barents Sea	Skern-Mauritzen *et al.* (2011)
			South Pacific	Great Barrier Reef	Smith *et al.* (2012)
			North Atlantic	Mid-western North Atlantic	Hamazaki (2002)
			Antarctic	65°W-55°W	Williams *et al.* (2006)

Table 4. Type of cetacean data and response variables used in species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to March 2015 are considered.

Reference	Cetacean data	Response variable
Ainley *et al.* (2012)	Opportunistic sighting	Presence only
Ballard *et al.* (2012)	Dedicated and opportunistic sighting	Presence only
Becker *et al.* (2012)	Dedicated sighting	Count
Beekmans *et al.* (2010)	Dedicated sighting	Count
Best *et al.* (2012)	Dedicated sighting	Presence/absence
Bombosch *et al.* (2014)	Opportunistic sighting	Presence only
Corkeron *et al.* (2011)	Opportunistic sighting	Count
Cotté *et al.* (2009)	Opportunistic sighting	Count
Dalla Rosa *et al.* (2012)	Opportunistic sighting	Count
Forney *et al.* (2012)	Dedicated sighting	Count
Friedlaender *et al.* (2011)	Opportunistic sighting	Presence only
Gregr (2011)	Catch	Presence only
Gregr and Trites (2001)	Catch	Count
Hamazaki (2002)	Dedicated sighting	Presence/absence
Hedley *et al.* (1999)	Dedicated sighting	Count
Kaschner *et al.* (2006)	Published data	-
Keller *et al.* (2012)	Dedicated sighting	Count
Laidre *et al.* (2010)	Dedicated sighting	Presence only
Laran and Gannier (2008)	Opportunistic sighting	Presence/absence
Moses and Finn (1997)	Dedicated sighting	Presence/absence
Murase *et al.* (2013)	Dedicated sighting	Count
Murase *et al.* (2014)	Dedicated sighting	Count
Okamura (2001)	Dedicated sighting	Count
Panigada *et al.* (2008)	Opportunistic sighting	Presence/absence
Pardo *et al.* (2015)	Dedicated sighting	Count
Pendleton *et al.* (2012)	Dedicated sighting	Presence only
Sasaki *et al.* (2013)	Dedicated sighting	Presence/absence
Skern-Mauritzen *et al.* (2011)	Opportunistic sighting	Count
Skov *et al.* (2008)	Opportunistic sighting	Presence only
Smith *et al.* (2012)	Opportunistic sighting	Presence only
Torres *et al.* (2013)	Catch	Presence/absence
Vikingsson *et al.* (2015)	Dedicated sighting	Count
Wheeler *et al.* (2012)	Various (e.g. sighting and catch)	Presence only
Williams *et al.* (2006)	Opportunistic sighting	Count
Williams *et al.* (2011)	Dedicated sighting	Count
Williams *et al.* (2014)	Opportunistic sighting	Count

Table 5. Explanatory variables used in species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to March 2015 are considered. SST: sea surface temperature; SSH: sea surface height (including its anomaly); Chl: sea surface chlorophyll *a* concentrations; Current: sea surface current; SSS: sea surface salinity.

Reference	Latitude	Longitude	Year	Month	Distance to shore	Depth	Slope	Aspect	Bottom complexity	Distance from bottom terrain (e.g. shelf)	SST	SSH	Chl	Current	SSS
Ainley *et al.* (2012)	-	-	-	-	-	X	X	-	-	-	-	-	X	-	-
Ballard *et al.* (2012)	-	-	-	-	-	X	X	-	-	-	-	-	X	-	-
Becker *et al.* (2012)	-	-	-	-	-	X	X	-	-	-	X	-	-	-	-
Beekmans *et al.* (2010)	X	X	-	-	-	X	-	-	-	X	X	-	X	X	-
Best *et al.* (2012)	-	-	-	-	X	X	-	-	-	-	X	X	X	-	-
Bombosch *et al.* (2014)	-	-	-	-	-	X	X	-	-	-	X	X	X	-	-
Corkeron *et al.* (2011)	-	-	-	-	X	X	X		-	-	-	-	-	-	-
Cotté *et al.* (2009)	-	-	-	X	X	X	-	-	-	-	X	X	X	-	X
Dalla Rosa *et al.* (2012)	X	X	X	X	X	X	X		-	X	X	X	X	X	X
Forney *et al.* (2012)*	-	-	-	-	X	X	-	-	-	-	X	-	X	-	X
Forney *et al.* (2012)*	-	-	-	-	-	X	X	-	-	X	X	-	X	-	X
Friedlaender *et al.* (2011)	-	-	-	-	X	X	X	-	-	-	-	-	-	-	-
Gregr (2011)	-	-	-	-	-	X	-	-	-	-	X	-	-	X	-
Gregr and Trites (2001)	X	X	X	X	-	X	X	-	-	-	X	-	-	-	X
Hamazaki (2002)	-	-	-	-		X	X	-	-	-	X	-	-	-	-
Hedley *et al.* (1999)	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-
Kaschner *et al.* (2006)	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-
Keller *et al.* (2012)	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-
Laidre *et al.* (2010)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Laran and Gannier (2008)	-	-	-	-	-	X	-	-	-	X	X	-	X	-	-
Moses and Finn (1997)	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-
Murase *et al.* (2013)	X	X	-	-	-	X	-	-	-	X	-	-	-	-	-
Murase *et al.* (2014)	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-
Okamura (2001)	X	X	X	X	-	-	-	-	-	-	X	-	-	-	-
Panigada *et al.* (2008)	X	X	X	-	X	X	X	-	-	-	X	-	X	-	-
Pardo *et al.* (2015)	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-
Pendleton *et al.* (2012)	-	-	-	-	-	X	-	-	-	-	X	-	X	-	-
Sasaki *et al.* (2013)	-	-	-	-	-	X	-	-	-	-	X	X	X	-	-
Skern-Mauritzen *et al.* (2011)	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-
Skov *et al.* (2008)	-	-	-	-	-	X	X	-	X	-	-	X	-	X	-
Smith *et al.* (2012)	-	-	-	-	X	X	X		-	X	X	-	-	-	-
Torres *et al.* (2013)	-	-	-	-	-	X	X	-	-	X	X	-	X	-	-
Vikingsson *et al.* (2015)	-	-	-	-	X	X	X	X	-	X	X	X	-	-	-
Wheeler *et al.* (2012)	-	-	-	-	X	X	X	-	-	-	X	-	X	-	-
Williams *et al.* (2006)	X	X	-	-	X	X	-	-	-	-	-	-	-	-	-
Williams *et al.* (2011)	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-
Williams *et al.* (2014)	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-

* Same paper but different covariates were applied different areas

Table 5. (continue)

Reference	Subsurface temperature	Subsurface salinity	Mixed layer depth	Thermal front	Oceanic front	Water mass	Distance from sea ice	Sea ice cover	Sea ice concentration	Prey	Wind speed /Beaufort	Length of day
Ainley *et al.* (2012)	-	-	-	-	X	-	-	X	-	-	-	-
Ballard *et al.* (2012)	-	-	-	-	X	X	-	X	-	-	-	-
Becker *et al.* (2012)	-	-	-	-	-	-	-	-	-	-	X	-
Beekmans *et al.* (2010)	-	-	-	-	X	-	X	-	-	-	-	-
Best *et al.* (2012)	-	-	-	-	-	-	-	-	-	-	-	-
Bombosch *et al.* (2014)	-	-	-	-	-	-	X	-	X	-	-	X
Corkeron *et al.* (2011)	-	-	-	-	-	-	-	-	-	-	-	-
Cotté *et al.* (2009)	-	-	-	-	-	-	-	-	-	-	-	-
Dalla Rosa *et al.* (2012)	X	X	-	X	-	-	-	-	-	-	-	-
Forney *et al.* (2012)	-	-	X	-	-	-	-	-	-	-	X	-
Forney *et al.* (2012)	-	-	X	-	-	-	-	-	-	-	X	-
Friedlaender *et al.* (2011)	X	-	-	-	-	-	X	-	-	X	-	-
Gregr (2011)	X	-		-	-	-		-	-	-		-
Gregr and Trites (2001)	-	-	-	-	-	-	-	-	-	-	-	-
Hamazaki (2002)	-	-	-	X	-	-	-	-	-	-	-	-
Hedley *et al.* (1999)	-	-	-	-	-	-	X	-	-	-	-	-
Kaschner *et al.* (2006)	-	-	-	-	-	-	X	-	-	-	-	-
Keller *et al.* (2012)	-	-	-	-	-	-	-	-	-	-	X	-
Laidre *et al.* (2010)	-	-	-	-	-	-	-	-	-	-	-	-
Laran and Gannier (2008)	-	-	-	-	-	-	-	-	-	-	-	-
Moses and Finn (1997)	-	-	-	-	-	-	-	-	-	-	-	-
Murase *et al.* (2013)	X	X	-	-	-	-	-	-	-	X	-	-
Murase *et al.* (2014)	-	-	-	-	X	-	-	-	-	-	-	-
Okamura (2001)	-	-	-	-	-	-	-	-	-	-	-	-
Panigada *et al.* (2008)	-	-	-	-	-	-	-	-	-	-	-	-
Pardo *et al.* (2015)	-	-	-	-	-	-	-	-	-	-	-	-
Pendleton *et al.* (2012)	-	-	-	-	-	-	-	-	-	X	-	-
Sasaki *et al.* (2013)	-	-	-	-	-	-	-	-	-	-	-	-
Skern-Mauritzen *et al.* (2011)	X	-	-	-	-	-	-	-	-	-	-	-
Skov *et al.* (2008)	X	X	-	-	-	-	-	-	-	-	-	-
Smith *et al.* (2012)	-	-	-	-	-	-	-	-	-	-	-	-
Torres *et al.* (2013)	X	-	X	-	-	-	-	-	-	-	-	-
Vikingsson *et al.* (2015)	-	-	-	-	-	-	-	-	-	-	-	-
Wheeler *et al.* (2012)	-	-	-	-	-	-	X	-	-	-	-	-
Williams *et al.* (2006)	-	-	-	-	-	-	-	-	-	-	-	-
Williams *et al.* (2011)	-	-	-	-	-	-	-	-	-	-	-	-
Williams *et al.* (2014)	-	-	-	-	-	-	X	-	X	-	-	-

Table 6. Summary of species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to March 2015 are considered.

Reference	Resolution	Model	Error distribution	Model evaluation	Spatial autocorrelation	Collinearity	Density surface estimation
Ainley *et al.* (2012)	5 km	MaxEnt	-	AUC	-	-	MaxEnt
Ballard *et al.* (2012)	5 km	MaxEnt	-	AUC	-	-	MaXent
Becker *et al.* (2012)	5 km	GAM	Quasi-Poisson	Visual comparison Comparison with DISTANCE estimate Spearman rank correlation test	-	-	IDW
Beekmans *et al.* (2010)	0.2°	GAM	Tweedie, Quasi-Poisson	-	-	-	GAM
Best *et al.* (2012)	10 km	GAM	Quasi-binomial	AUC	-	-	GAM
Bombosch *et al.* (2014)	0.25°	MaxEnt	-	AUC	-	-	MaxEnt
Corkeron *et al.* (2011)	0.1°	GLM	Quasi-Poisson	-	SEVM	-	GLM
Cotté *et al.* (2009)	NA	GAM	Gamma	-	-	-	GAM
Dalla Rosa *et al.* (2012)	4.63 km	GAM	Quasi-Poisson	-	Variogram	-	GAM
Forney *et al.* (2012)	10 km	GAM	Quasi-Poisson	-	-	-	IDW
Forney *et al.* (2012)	5 km	GAM	Quasi-Poisson	-	-	-	IDW
Friedlaender *et al.* (2011)	1 km	MaxEnt	-	AUC	-	-	MAxEnt
Gregr (2011)	50 km	MaxEnt	-	AUC		-	MaxEnt
Gregr and Trites (2001)	10 km	GLM	Poisson	Crossvalidation Classification tables	-	-	GLM
Hamazaki (2002)	10'	Logistic regression	Binomial	-	-	-	Logistic regression
Hedley *et al.* (1999)	5 n.miles	GAM	Poisson	Comparison with DISTANCE estimate	-	-	GAM
Kaschner *et al.* (2006)	0.5°	RES	-	Comparison with actual data	-	-	RES
Keller *et al.* (2012)	4 km	GAM	Poisson	Bootstrap	GLMM	-	GAM
Laidre *et al.* (2010)	2 km	Kernel method	-	-	-	-	Kernel method
Laran and Gannier (2008)	10 n.miles	Logistic regression	Binomial	AUC	-	-	Logistic regression
Moses and Finn (1997)	10'	Logistic regression	Binomial	Visual comparison	-	-	Logistic regression
Murase *et al.* (2013)	10 km	GAM	Poisson	Comparison with DISTANCE estimate	-	-	GAM
Murase *et al.* (2014)	30 km	GAM	Tweedie	Comparison with DISTANCE estimate	-	VIF	GAM
Okamura (2001)	1°	GAM	Poisson	-	-	-	GAM
Panigada *et al.* (2008)	NA	GAM	Binomial	-	-	-	GAM
Panigada *et al.* (2008)	NA	CART	-	-	-	-	CART
Pardo *et al.* (2015)	1/3°	HBM	-	-	-	-	Hierarchical Bayesian
Pendleton *et al.* (2012)	1 km	MaxEnt	-	AUC	-	-	MaxEnt
Sasaki *et al.* (2013)	4 km	GLM	Binomial	-	Moran's I	-	GLM
Skern-Mauritzen *et al.* (2011)	50 km	GAMM	Quasi-Poisson	-	-	VIF	GAMM
Skov *et al.* (2008)	1 km	ENFA	-	-	-	-	ENFA
Smith *et al.* (2012)	4.8 km	MaxEnt	-	AUC	-	-	MaxEnt
Torres *et al.* (2013)	25 km	BRT	-	AUC	-	-	BRT
Vikingsson Víkingsson and Heide-Jørgensen (2015)*et al.* (2015)	0.5°	GAM	Negative binomial	-	Variogram	Pearson Correlation coefficient	Kriging
Wheeler *et al.* (2012)	10 km	ENFA	-	Jack-knife	-	-	ENFA
Williams *et al.* (2006)	5 km	GAM	Quasi-Poisson	-	-	-	GAM
Williams *et al.* (2011)	20 n.miles	GAM	Tweedie	-	-	-	GAM
Williams *et al.* (2014)	6.25 km	GAM	Tweedie	-	-	-	GAM

Table 7. Summary of generalized additive models (GAMs) applied to baleen whales as species distribution models (SDMs). Papers published in scientific journals from December 1997 to March 2015 are considered.

Authors	Software	Version	Package	Version	Error distribution	Smoother	Dimension parameter (k)	gamma	Model selection
Becker *et al.* (2012)	S-PLUS	6.1 Release 1	-	-	Quasi-Poisson	NA	NA	NA	AIC
Beekmans *et al.* (2010)	R	NA	mgcv	1.5-5	Tweedie Quasi-Poisson Poison	Isotropic/ tensor product	NA	1.4	GCV
Best *et al.* (2012)	R	NA	mgcv	NA	Quasi-binomial	Thin-plate	NA	1.4	GCV
Cotté *et al.* (2009)	R	NA	mgcv	NA	Gamma	NA	NA	NA	GCV
Dalla Rosa *et al.* (2012)	R	NA	mgcv	1.4-1	Quasi-Poisson		8	1.4	GCV
Forney *et al.* (2012)	S-PLUS R R	NA 2.6.2 2.6.2	- gam mgcv	- NA 1.3-29	NA	Cubic spline Cubic spline Cubic and thin-plate spline	3 3 NA	1.0 & 1.4	AIC AIC GCV
Hedley *et al.* (1999)	NA	NA	NA	NA	Overdispersed-Poisson	Cubic spline	2, 4, 8	NA	AIC
Keller *et al.* (2012)	NA	NA	NA	NA	Poisson	NA	2	NA	AIC
Murase *et al.* (2013)	R	2.12.1	mgcv	1.7-2	Poisson	NA	NA	NA	GCV
Murase *et al.* (2014)	R	3.0.2	mgcv	1.7-26	Tweedie	NA	NA	NA	GCV
Okamura *et al.* (2001)	S-PLUS	NA	-	-	Poisson	NA	NA	NA	BIC
Vikingsson *et al.* (2015)	R	NA	mgcv	1.8-1	Negative binomial	Cubicspline	NA	NA	GCV
Williams *et al.* (2006)	R	NA	mgcv	NA	Overdispersed-Poisson	NA	NA	NA	GCV
Williams *et al.* (2014)	R	NA	mgcv	NA	Tweedie	NA	NA	NA	REML
Williams *et al.* (2011)	R	NA	mgcv	NA	Tweedie	Soap-film	NA	NA	NA

*Indicated by discussions with authors.